understanding factors :

(Y) Balance ~ income (X_1)

↳ purpose: prediction

↳ use lm(·)

Balance ~ income + Region ↗ Factor = { East, south, west

(how does R encode this?

↳ K categories → K-1 dummy variables

$$\text{Balance} = \hat{B_0} + \hat{B_1}\,\text{income} + \hat{B_2}\,X_{south} + \hat{B_3}\,X_{west}$$

Credit

	Balance (Y_i)		income (X_1)	Region (X_2)
Y_1	14.891	X_{11}	333	South
Y_2	106.025	X_{12}	903	West
Y_3	104.593	X_{13}	580	West
				⋮

X_{ii}

X_{south} = { 1 if S, 0 if o/w

X_{west} = { 1 if S, 0 if o/w

↳ $\hat{B} = \begin{pmatrix} \hat{B_0} \\ \hat{B_1} \\ \hat{B_2} \\ \hat{B_3} \end{pmatrix} = (X^T X)^{-1} X^T Y$

Y = Balance ; $X = \begin{pmatrix} 1 & X_1 & X_{south} & X_{west} \\ 1 & \vdots & & \\ 1 & & & \\ \vdots & & & \end{pmatrix}$